Social Psychology Quarterly
2017, Vol. 80(1) 1–19

DOI: 10.1177/0190272516680842
journals.sagepub.com/home/spq

Online Field Experiments: Studying Social Interactions in Context

Paolo Parigi[1], Jessica J. Santana[1], and Karen S. Cook[1]

Abstract

Thanks to the Internet and the related availability of "Big Data," social interactions and their environmental context can now be studied experimentally. In this article, we discuss a methodology that we term the online field experiment *to differentiate it from more traditional lab-based experimental designs. We explain how this experimental method can be used to capture theoretically relevant environmental conditions while also maximizing the researcher's control over the treatment(s) of interest. We argue that this methodology is particularly well suited for social psychology because of its focus on social interactions and the factors that influence the nature and structure of these interactions. We provide one detailed example of an online field experiment used to investigate the impact of the sharing economy on trust behavior. We argue that we are fundamentally living in a new social world in which the Internet mediates a growing number of our social interactions. These highly prevalent forms of social interaction create opportunities for the development of new research designs that allow us to advance our theories of social interaction and social structure with new data sources.*

Keywords

online field experiment, Big Data, Internet studies, experimental social psychology, quantitative methodology

Social interactions are increasingly digitized. Thanks to the Internet and the availability of "Big Data," large-scale social interactions can now be studied experimentally without the need to bring people into onsite laboratories. Many of these forms of social interaction are best studied in the environment in which they occur in order to fully understand the factors that influence them and their dynamics over time. This is made possible by the availability of large data archives as well as organizations that collect individual-level data on a scale previously unimaginable.

In this article, we discuss a relatively new methodology, the *online field experiment*, that takes advantage of Big Data and predictive algorithms to capture

[1]Stanford University, Stanford, CA, USA

Corresponding Author:
Paolo Parigi, Institute for Research in the Social Sciences, Stanford University, 30 Alta Road, Stanford, CA 94305, USA.
Email: pparigi@stanford.edu

complex environmental treatments. The goal of the online field experiment is not to generalize findings from a laboratory experiment or investigate the effects of a phenomenon. Instead, the goal is to test hypotheses derived from a theory that incorporates "treatment complexity" in cases in which the social interaction under study cannot be isolated from its contextual environment without loss of meaning. Treatment complexity implies more than the context providing meaning to social interactions while the experimenter manipulates conditions under which these interactions occur. As we will explain, a treatment is complex when the meaning of the social interaction being investigated is fundamentally connected to the context in which the social interactions take place (Ashmore, Deaux, and McLaughlin-Volpe 2004).

For example, lab experiments of trust interactions typically use some version of an investment game (Berg, Dickhaut, and McCabe 1995) to measure "thin" trust, the type of trust that is generated between strangers with no prior or anticipated future relationship. This is not the type of "thick" trust that is most often embedded in social relationships and very difficult to isolate for study in a laboratory (Cook 2005). Online field experiments, in contrast to traditional or online lab experiments, can capture the context in which these types of social relationships occur.

The purpose of the traditional laboratory experiment, which often involves testing hypotheses derived from a theory rather than the investigation of a specific "effect" in the laboratory (e.g., the classic bystander intervention studies or the Asch conformity studies: Asch 1951; Darley and Latane 1968; Latane and Darley 1968), is not to directly replicate a phenomenon (see the distinction between theory-driven and effect-oriented experiments in Zelditch 2014). Neither is the primary goal of an online field experiment to capture the entire set of conditions surrounding the key phenomenon of interest. Instead, the purpose of most experiments in general is to "construct and test theories" by creating "theoretically relevant aspects of social situations under controlled conditions" (Zelditch 1969:530).

Theoretical laboratory experiments are necessarily abstractions of theoretical concepts and not meant to reflect the reality of concrete instantiations of those concepts (Zelditch 1969:530). Thus, theoretical experiments are not designed to replicate natural "armies in the lab"—they are designed to simulate the aspect of armies (e.g., an authority relationship) that is relevant to the theory in question (e.g., a theory of the exercise of legitimate power). This article addresses theoretical experiments in which "complexity" (Zelditch 1969:533), or more precisely, the interdependency between context and the relevant social interactions, is a critical aspect of the theory under question. We argue that theories involving such complexity have traditionally been outside the scope of laboratory experiments because of prior logistical restrictions (Zelditch 1969). Researchers now acknowledge the value of Internet technology in accessing larger and more dispersed participant pools as well as vast amounts of data on various types of social and economic interactions. Our article emphasizes the fact that the "Big Data" currently being collected in a wide variety of online contexts, such as sharing economy websites, expands our capabilities for including relevant aspects of the context in which significant social and economic interactions occur in our experimental designs.

We use the term *Big Data* to mean the *in-context* documentation of individual participant behavior, preferences, and attributes (McFarland, Lewis, and Goldberg 2015). Examples of Big Data include documentation of health and exercise activity via Internet-enabled bicycle sensors, mobile phone call and SMS records, emoticon reactions to status updates on Facebook, career histories on LinkedIn, credit card transactions, online movie rental preferences, and presentations of identities via website profiles, among others. The key aspect of Big Data is that it is collected at the individual level and incorporates social interactions between individuals in a variety of contexts. While other definitions of Big Data are available, the capacity of capturing treatment complexity within an experimental framework rests on the fact that Big Data now include individual-level records of millions of social and economic interactions and thick descriptions of the contexts in which they occur. Online field experiments capture treatment complexity using Big Data.

The online field experiment method we describe in this article emphasizes the idea that the Internet is not just another mechanism for recruiting more subjects (see i.e., Amazon Mechanical Turk). Rather, the Internet is a place where many platforms facilitate and collect data on the social and economic interactions (online and offline) of their members. Big Data routinely collected on these interactions make the overall social environment more measurable. Researchers, for example, can work with platforms hosting various online communities to randomly select a sample of study participants and then use the Big Data these platforms collect on the selected individuals and the contexts of their interactions. Researchers can use the algorithms that many of these online communities use to predict whether a participant will be exposed to a treatment of interest. In fact, the prediction of future experiences can become part of how the sample is drawn in the first place. By studying participant behavior in an online field setting, researchers are able to preserve what we have called treatment complexity while testing theoretically driven hypotheses.

In the following sections, we describe online field experiments and the situations in which they are most valuable. We further define treatment complexity and explain how Big Data enable online field experiments to account for complex treatments. We then provide the details of an example of an online field experiment that we used to investigate the impact of the sharing economy on trust behavior. Our study suggests that in a world where the Internet increasingly mediates our social interactions, experimentation can include theoretically relevant treatments while studying social interactions in situ, providing access to much larger samples and newer types of data than has previously been possible in physical laboratories.

ONLINE FIELD EXPERIMENTS

Online field experiments are distinct from online lab experiments (cf. Centola 2010; DeVoe and Pfeffer 2007; Goldstein and Hays 2011; Hahl and Zuckerman 2014; Salganik, Dodds, and Watts 2006) in that they are built into the context of the online community under study. An online lab experiment creates a kind of controlled laboratory situation in the online environment. While random assignment is possible using online lab experiments, capturing the contextual environment of the social interaction under study requires a field design instead. Salganik et al. (2006) and Centola (2010), for example, employ random assignment, but their research also simulates the study environment rather than relying on a preexisting environment in which participants were already embedded. Their goal is to create a world that is similar to the environment they wish to investigate in the ways they think are most relevant for testing their hypotheses, but in doing so, they have to simulate that environment rather than study the participants' actual engagement in such an environment.

As in natural and other quasi experiments, online field experiments do not allow for a direct manipulation of the treatment (Campbell and Stanley 1963; Cook and Campbell 1979). Online field experiments need not assign subjects at random to either control or treatment conditions. Instead, online field experiments select a random sample of a website's (or other online community's) population for participation, divide participants into treatment and control groups using Big Data and predictive algorithms to predict whether the participant will be exposed to the treatment or not, and then observe or measure the participants' outcomes. The experimenter does not assign people to treatment or nontreatment conditions. Individuals self-select into conditions.

Our version of online field experiments uses Big Data to leverage the selection bias inherent in this type of design. Big Data are used for understanding the people in the sample that are participating in the study compared to those that are not and for reducing the impact of attrition, namely, of individuals that drop out of the study. Big Data are useful in linking an individual participant's experimental data with their behavioral and demographic data observed (and typically recorded) in the online community. Online field experiments do not rely on Big Data for generalization to a broader population but rather to control for different types of biases mentioned previously. Because online field experiments capture both the context and the entire interaction, from beginning to end, within the online community, researchers can more readily rely on data collected online to analyze many potential intervening factors that affect the treatment(s) of interest.

Online field experiments are necessary when the environment generates a level of treatment complexity that, while not reproducible in a traditional lab, is effectively captured using the data collected by the online community of reference. Like traditional field experiments (cf. Baldassarri and Grossman 2013; Correll, Bendard, and Paik 2007; Pager, Western, and Bonikowski 2009; Shepherd and Paluck 2015; Webster and Sell 2014), online field experiments permit some of the relevant complexity that defines the social phenomena of interest. When the "field" is an online community, that is, when technology mediates several aspects of the social interaction (including its initiation and meaning), then the context in which social interactions take place becomes bytes of information stored in the servers of a company that the apt researcher can harvest. In the next section, we define treatment complexity more fully and explain how online field experiments are well situated for research involving such complexity.

Treatment Complexity

The theoretical complexity of a treatment such as community membership means that the social interaction under study cannot be isolated from its contextual environment without loss of meaning (Ashmore et al. 2004). This goes beyond the argument that context provides the raw material for what happens during a social interaction and the relevant identities that get enacted (e.g., the context provides the stereotypes in research about stereotypes). We argue that context often directly impacts and mediates social interactions. *Treatment complexity* describes the degree to which a treatment entails embeddedness of the social interaction in its environment. Treatment complexity does not imply statistical interactions across multiple factors (though that may occur in some settings). A treatment is complex when the social interaction cannot be isolated from the context in which it happens without losing or significantly altering the overall effect of the experience. A complex treatment is one in which the interactions between participants are intrinsically shaped by the context in which they take place. The overall environment, composed of interactions embedded in a context, sets the conditions under which participants act and react. Moreover, participants engage directly with the environment, interpreting and altering the environment through their interactions (Page 2015).

This approach suggests that in a world where social behavior is increasingly mediated by technology, experimentation can include greater levels of treatment complexity than has previously been thought possible in physical laboratories. Technology allows for the collection of detailed data on social interactions that can later be used for (re)creating conditions similar to that of a traditional field experiment. The complex patterns of social interactions that defined the treatment can be accounted for and analyzed in an online field experiment. The methodology we are proposing is based on the idea that the power of Big Data is not so much in its size (the census has existed for decades) but in its depth—the amount of data collected on the social interactions between participants operating in the given context.

An implication of the social world becoming more digital is the availability of data on these types of social and economic interactions and the contexts in which they occur that were previously inaccessible. For example, data accumulated by users of traveling websites in the sharing economy can clarify the impact that sharing a room with a stranger has on signals of trustworthiness (e.g., similarity or reputation). Here, the abstract concept instantiated by "sharing a room with a stranger" or "loaning a car to a stranger" is "experience in the sharing economy." The sharing economy, like the "gig" economy, is a theoretical construct with implications for theories of exchange and trust, among others. As these economies spread online, Internet technology and the capture of Big Data have exposed researchers to the interconnections between context and social interactions. Such interdependency cannot be distilled into a traditional laboratory environment without significantly altering its meaning. Instead, the exposed interdependency between context and social interactions can now be studied with the online field experimental approach. With the expansion of the Internet into more domains of life, researchers can employ experimental designs without neglecting the complexity of these interactions embedded in the context in which they occur because we have precise, individual-level data on the people entering into these interactions (e.g., reputations, number of prior experiences, tone of the emails exchanged prior to meeting) and the context (e.g., location of the shared room, characteristics of the neighborhood, privacy of the room, etc.).

In another example, an online field experiment may be the preferred approach if the researcher seeks to capture the role of multiplex online relationships. Multiplex relationships are those occurring across multiple dimensions or contexts of interaction (Santana, Hoover, and Vengadasubbu 2017). A person might interact with another person in the contexts of work, recreation, and community service, for example. Experimentally manipulating the multiplexity of a relationship is extremely challenging. In a laboratory, the best designs might only achieve a proximate and relatively weak measure of multiplexity: such a design might assign a treatment condition of interaction with the same partner across diverse task domains, for example, while a control group only interacts in a single domain. In the field, however, researchers can directly measure multiplexity, which is intricately bound up with the environment in which the interactions occur. This dimension of complex social interactions cannot easily be abstracted or replicated in the lab. Interaction with the same person at work, in the gym, and at a volunteer event is starkly different from interaction with a stranger in a lab to solve a puzzle, write a poem, and build a model airplane. The social context is missing from the latter. Finally, if technology is fundamentally intertwined with the environment in which the multiplex social interactions take place, an online field experiment is a good way to capture the effects of these interactions or measure related behavioral changes. Online studies of dating, friendship, or other types of social exchange, for example, can account for the complexity of treatments like anonymous dating, exchange commitment, or distrust using rich participant data collected on websites like OkCupid, Facebook, or Last.fm.

In the case of online field experiments, the Internet is not used simply as another mechanism for recruiting more subjects. More importantly, the Internet mediates our interactions and provides them with meaning. This does not mean that online field experiments focus only on the analysis of *online* social interactions. Instead, it means that online field experiments can be used to leverage online *data* collected on social interactions in whatever context they occur: online or offline. The large amounts of data routinely collected on such social interactions and the context in which they occur thus make the effects of complex treatments measurable.

EXISTING WORK RELATED TO ONLINE FIELD EXPERIMENTS

Experiments conducted using computers are not a new phenomenon. Precursors to the online field experiment have been in use by computer scientists and social scientists alike for decades. The field of human-computer interaction, with foundations in WWII-era cybernetic experimentation (Hauben and Hauben 1998; Wiener 1948), quickly adopted applied online field experiments with the goal of improving the design of computer interfaces. As online communities grew in population and diversity, social scientists began exploring the rich data sets these sites generated on social behavior. Early forms of online field experiments involved emailing a randomly selected subset of website users with requests to engage in a new feature of the website, similar to modern-day email marketing efforts (Chen and Konstan 2015). As the value of experimental data became known to web-based companies, websites began actively seeking social scientific insights into their users' behavior. While the methodology is not unique, new forms of collaboration between online community hosts and researchers allow those using online field experiment designs to take advantage of the availability of massive behavioral data sets and global field access, creating fertile ground for a new type of social research.

Online communities now cover an extensive range of the global population. The largest online community is the social network Facebook. Over one billion people worldwide use Facebook each day (Facebook Company Info 2016). The vitality of this network permits deeper study of the dynamics of social and economic interactions both online and offline. In a 2012 *Nature* study, for example, social scientists from University of California, San Diego collaborated with Facebook employees to explore social influence and political mobilization (Bond et al. 2012). The researchers experimentally manipulated social influence. They randomly assigned users to one of three groups. In the "social message" condition, respondents were encouraged to vote by showing them a subset of their friends who had voted. In the second condition, an "informational message" was used to encourage voting but without displaying the subset of friends who had voted. The control group received no message at all encouraging them to vote. Not only did the researchers learn more about the distinct effects of tie strength, they also found that online social influence affected offline voting behavior *more than* online behavior did.

Online behavior includes a diverse array of social phenomena. Even interactions as intimate as courtship are observable in online settings. Websites like OkCupid have been collecting data on dating habits since at least 2009 (see OkTrends.OkCupid.com). Such data are not limited to U.S. communities. Ong and Wang (2015), for example, used data from one of the largest online dating websites in China to explore the effect of income on mate attraction. Like offline audit studies, they created a set of fake profiles using data from inactive profiles on another online dating website. Income and other attributes were systematically assigned to the fake profiles. The researchers then measured the effect of income on the number of visits to a fake profile. These new data on the determinants of social affiliation are difficult to obtain in any other way.

Outside of academia, many web-based companies use online experiments for capturing behavioral changes among their users. The simplest and most commonly used form is the A/B test, in which the researcher randomly assigns half of the participants to a test condition and the other half to a control condition (Kohavi and Longbotham forthcoming). In most cases, these industry-based researchers wish to know how a product change will influence customers' site utilization or purchasing behavior. By revealing the change in product (or process) to only half of their users, researchers are able to quantify the average treatment effect of the design change on behaviors of interest. This is mainly applied research, but the same platforms can be used for more theoretically motivated research efforts.

Unlike industry-based A/B testing and more similar to the online experiments previously described, the online field experiment is designed with the goal of detecting and testing theoretical insights. In support of the "army in a laboratory" perspective (Zelditch 1969), the goal of the online field experiment is to test the theoretically relevant aspects of the complex interactions that are expected to produce the observed behavioral changes in the field. The method we describe in this article does not suggest attempting to detect changes in behavior without a theoretical hypothesis informing the collection of data. The details of implementing an online field experiment are described in the next section.

RESEARCH DESIGN FOR AN ONLINE FIELD EXPERIMENT

There are three important components of online field experiment research design: collaboration with online platforms, recruitment of participants involved with the online community of interest, and retention of participants regardless of their probability of compliance with the treatment. We describe each of these components and how to address them in the following.

In designing an online field experiment, we recommend that researchers first identify the boundaries of the community of interest. This community is, broadly speaking, the environment within which the social interactions occur in a context. For example, an online community could include players of an online game, traders in an online market, contributors to an online software project, or followers of an online persona. After a community has been identified, the researcher needs to contact representatives of the online community platform in charge of maintaining the community. Collaboration with the administrators (or owners) of the platform is fundamental for engaging participants of the community and testing the response bias of those who do participate. The goal of an online field experiment is not to maximize external validity by obtaining a representative sample of some generic "population." After the collaboration is arranged, the researcher or the administrators of the platform (or both) invite participants to engage with the study using emails, ads placed on the community platform, forum posts, or other relevant media.

Collaboration with those who manage the platform is also fundamental for dividing participants into the treatment and "nontreatment" conditions. Treatment, in this case, is not participation in the online community but rather some behavior or attitude occurring *within* the community (e.g., hosting travelers in one's home or producing material for an online website). Treatment assignment may appear impossible at first because it seemingly requires that the researcher have the capacity to predict the future at the time of the creation of the study (or in deciding whom to invite to participate). However, online platforms routinely use predictive models and machine learning to predict the future behavior of their participants. That is, online field experiments directly harness the power of Big Data by incorporating predictions about future behavior. Because this classification is not random and is instead based on predictive models and because participants that do not experience treatment during the period of study are still expected to experience the treatment later, we use the term *nontreated* in contrast to *control*.

An online field experiment can include within-subject (before and after) or between-subject (posttest) designs, among others. After recruitment, participants in the before-after design are measured on the outcome variable of interest, such as general attitudes toward trust. This measurement takes place in the online lab and may consist of the participant playing a behavioral game. After treatment or nontreatment occurs in the field, participants are invited back to the online lab and measured a second time. In the posttest design, measurement of the treated and nontreated group outcomes occurs at a fixed point in time after all participants expected to receive the treatment have actually received it. The time between the before and after measures needs to be carefully calculated on the basis of the measurement tool the experimenter plans to use. In the case of a behavioral game, in a before-and-after design, the experimenter needs to consider that a participant can learn how to play the game after the first time and use this strategy the second time he or she plays the game.

Obtaining participants for an online field experiment is much more challenging than recruiting students or Amazon Mechanical Turk volunteers. A researcher must first convince those who manage or control the online platform to collaborate with him or her and then must persuade prospective volunteers to participate in the study. In a before-after design, retention of participants is of paramount concern and a significant threat to the internal validity of the experiment. As described in the following, poor retention can introduce selection bias into the study. The standard solution for this problem is to offer incentives, usually in the form of monetary rewards. For an online field experiment, however, this may be cost prohibitive because of the volume of potential participants: Even just offering $5 to each of 9,000 participants would require a minimum budget of $45,000.

In an online field experiment that we ran, we solved this problem in two ways. First, we made prizes for participation more substantial (i.e., $100 gift card), though available only to a few; and second, we created the illusion that participants were playing a game in two phases and that they needed to come back to Phase 2 in order to know their scores in Phase 1. To be more specific, the allocation of gift cards was not a lottery but was in itself a game, where the chances of winning a gift card depended on the participant's winnings in the game. Overall, this strategy yielded a retention rate of about 65 percent.

Incentivizing participation through monetary rewards invariably can attract malfeasants who wish to circumvent the study to access the reward. Such people can, for example, program "bots" to automatically register for and collect study rewards multiple times without any human participation. Online field experiments thus require investing in a robust enough security system to prevent easy access to the reward without valid participation as well as the careful protection of participants' personally identifying information. More security, however, requires more effort by participants to prove their identity. A researcher who plans to use the online field experiment design needs to consider the tradeoff in costs and benefits between attracting and retaining more participants and the security of the online lab.

SOURCES OF BIAS IN ONLINE FIELD EXPERIMENTS

In a perfectly randomized experiment, subjects are assigned to treatment or control conditions at random. One could imagine this assignment to be the result of flipping a coin, for example. This makes the subjects exchangeable in the sense that their outcome measurements are, on average, not the result of individual differences. This exchangeability of subjects across conditions thus neutralizes the effect of confounding factors by making the effect of being treated (T = 1) the same as if everybody had been treated (t = 1).[1] Formally,

$$P[Y^{t=1}=y]=P[Y=y|T=1]. \qquad (1)$$

Because the treatment in an online field experiment involves a high degree of complexity, there is a high risk that recruitment into the study and retention in a before-and-after design are not independent of the participant's compliance with the treatment. A participant complies with a treatment when they receive the treatment, or nontreatment, that they were originally predicted to experience. When noncompliance is randomly distributed across treatment groups, the study results are not significantly affected. However, noncompliance can bias study results if an unobserved variable, such as household income, influences participants to sign up or prematurely drop out of the study. The more complex the treatment is, the higher the risk of this selection bias. For example, a participant's level of engagement with a specific community is vulnerable to influence

[1]A more precise terminology for explaining Equation 1 is that the conditional distribution of the outcome being treated is the same as the unconditional distribution of the outcome had everybody been treated. We thank Xiaolu Wang for clarifying this.

from a variety of factors, including the demographic composition of the community, activity level of other community members, and reliability of the community platform service. Through collaboration with online community host organizations, researchers can reduce the risk of unobserved variable bias and use techniques such as inverse probability weighting to predict a participant's probability of compliance.

The online field experiment addresses compliance bias using inverse probability weights that randomize the likelihood of dropout between waves with respect to observed characteristics of the participants. This makes the distributions of the outcome the same as in the unweighted original population. Formally,

$$W = \frac{Pr[C=0]}{Pr[C=0|L]}, \tag{2}$$

where C = whether a respondent dropped out between waves (0 = no; 1 = yes), and L = the set of independent variables measured at the beginning of the wave. This weighting scheme randomizes a participant's likelihood of dropping out of the study with respect to known characteristics (L). The matrix of L variables is constructed using Big Data, namely, the data that the platform collects at the individual level on its users. Selection remains but not the bias. Santana and Parigi (2015) estimated W using a logit model and applied the predicted probabilities in a weighted regression in a study of the effects of engagement in the sharing economy on changes in risk attitudes.

The other source of bias facing the online field experiment is selection into recruitment. This source of bias affects both designs but is more problematic for the post-test design since subjects act as their own control group in the before-after design.

Selection into recruitment is not the same as selection into treatment or control conditions in traditional experiments. Subjects in an online field experiment self-select into treatment or nontreatment categories, and recruitment mainly serves the purpose of measuring the outcome of interest. Because recruitment does not yet involve "treatment," there is less concern that participants will drop out *because of the treatment*.

The online field experiment addresses selection into recruitment by creating a random sample of users of the platform that then receive an invitation to participate in the study. This makes the invitation to participate in the study, namely, recruitment, independent of experiencing the treatment. The creation of a random sample of invitees is a key aspect of the online field experiment for two reasons. First, if the group of invitees that agree to participate in the study, namely, the participants, is large (e.g., a 15 percent response rate), then potential selection bias is greatly reduced. Second, the bias introduced by self-selection into participation can readily be measured and accounted for by comparing the characteristics of those who received the invitation and did not participate against those who decide to participate.

As we will illustrate via an example in the following, the random sample of invitees is constructed separately for treatment and nontreatment groups using predictive algorithms. These algorithms can predict who is going to experience the treatment and who is not before that treatment (in the field, that is) actually occurs. Big Data are also used to create balance between the two groups such that an equal proportion of, say, women in the treatment and nontreatment groups receive invitations to participate. Theoretical or substantive considerations can be used to create a stratified random sample. For example, in a study we conducted with Airbnb, we presumed that hosts and guests are different with respect to the amount of trust required for participating in the platform. We therefore stratified the random sample with respect to participants' host and guest status.

In sum, our implementation of an online field experiment allows us to directly control for both compliance and selection bias. The capacity to take bias into account in our estimates is where our research methodology differs the most from using Amazon Mechanical Turk or other online methods simply for recruiting subjects. In the next section, we provide an example of an online field experiment and illustrate our method for addressing the concerns of selective attrition and recruitment in an online field experiment.

AN EXAMPLE: TRUST AND REPUTATION IN THE SHARING ECONOMY

The sharing economy describes the growing ecosystem of providers and consumers of temporary access to products and services. At the helm of this economy are technology companies using the Internet with the goal of increasing the efficiency with which people connect to one another. Depending on the definition, surveys estimate that sharing economy participants make up 19 percent to 40 percent of the U.S. population (Owyang, Samuel, and Grenville 2014; PricewaterhouseCoopers 2015). The majority of these participants are consumers of the service, and only 7 percent provide a sharing economy service. According to a 2015 study, urban millennials, 18 to 24 years old, with household incomes from $50,000 to $75,000 are the most frequent users of such services (PricewaterhouseCoopers 2015).

In this section, we describe in broad terms an application of the online field experiment methodology to the question of how experience in the sharing economy affects how participants use information to place trust in strangers (the full study results will be presented in a separate publication). Note that our treatment variable is *experience*, not participation, in the sharing economy. Our dependent variable is trust. The "sharing economy" is a complex social system resulting from the dynamic intersection of identities, incentives, and structures. In many sharing economy marketplaces, members rely entirely on interpersonal trust as a form of currency. Thus, seemingly unrelated transactions, such as pet sitting and ridesharing, require that an individual trust an unknown stranger who is a member of the same community. In the eyes of the participants, membership in these communities has important effects. For instance, a recent survey found that 78 percent of the users in the sharing economy felt that their online interactions with people made them more open to the idea of sharing with strangers outside of the sharing economy (Visual.ly 2012). Studying companies within the sharing economy means studying not only the impact that technology has on individuals' behavior online but also the implications of such behavior for offline interactions and associations, potentially extending trust.

We recently implemented an online field experiment in collaboration with Airbnb, a home and room rental service that is hosted online and represents one community in the sharing economy. Our goal was to measure whether experience in the site changed the type of information Airbnb members used when deciding whom to trust (note that trust is the outcome variable in this case, not the treatment of interest). We contrasted two sources of information about unknown others: demographic characteristics (e.g., gender and age) and reputation acquired in the platform (e.g., the 5-star rating system).

Traveling using Airbnb was our complex treatment, and the amount of trust toward another unknown member of Airbnb was our outcome of interest. We asked: Does the reputation members acquire using Airbnb allow for the extension of interpersonal trust toward others with different sociodemographic characteristics? While the details of how we measured trust are not fundamental for the scope of this article, we provide a brief overview of these details before explaining the research design in greater detail.

To measure our outcome of interest (i.e., trust), participants in our online field experiment played a modified version of Berg et al.'s (1995) Investment Game. They were assigned a total of 100 points to invest in five different exchange partners. Whatever the participant decided to invest, that amount was multiplied by a factor of three and given to the recipients. By multiplying the investment by three, the strategic player can obtain the highest return by investing all 100 points in the partner most expected to return the full tripled amount. For example, the player may have decided to allocate his or her investment decisions evenly across the five recipients, namely, 100 / 5 = 20 points each, or may have decided to favor one recipient more than the others by giving, say, 40 points to that recipient, leaving the rest to be divided among the remaining four. If he or she chose, the participant could have decided to give all the points to one opponent and zero to everybody else or to keep all the points for himself or herself.

In this game, the recipients were not real people but simulated "robots," or actors that were constructed to represent the characteristics of a general "other" with certain attributes. Each of the five simulated actors was different from the participant who engaged in the study to create variation in the social distance between the participants and those to whom they responded. We determined distance by randomly combining the attributes of the simulated actors and then measuring the distance of each from the player. Thus, an actor with a few attributes in common with the player would have a smaller social distance from the player than an actor with no attributes in common and would have a larger social distance from the player than an actor with all attributes in common with the player. We employed one simulated actor that was close in social space to the player, one that was midway, and a third that was far away (increasing the social distance across the actors). The order in which the simulated actors appeared was randomized.

This setup created various measures of trust—overall points invested, points invested in the closest robot, points invested in the most distant robots, and so on. Finally, the study took place in two phases. In Phase 2, participants were matched with the same simulated actors as in Phase 1 and played the same game. A behavioral change in the defined measure of trust can be measured as the difference between investments in Phase 1 and 2.[2]

We describe in the following how we used Big Data and bias profiles to address recruitment bias in this study.

Dealing with Recruitment Bias

For this study, we collaborated with Airbnb to use Big Data collected on participants through the Airbnb platform. These data included demographics such as age and gender as well as behavioral

[2]Trust could be distinguished from risk aversion through the use of Holt and Laury's (2002) risk lottery game to measure participants' risk profiles. This measurement should be taken before and after treatment, however, as the treatment could influence the participant's threshold for risk (Santana and Parigi 2015).

data such as the number of times the participant hosted or was hosted by another Airbnb user, scheduled host or guest experiences, and temporal patterns of hosting or being hosted. We used these data to predict treatment and identify recruitment bias among participants.

A few weeks before starting Phase 1 (in September 2015), Airbnb created a random sample of about 120,000 users in the United States. The sample was stratified along two dimensions: (1) role as host or guest and (2) level of prior experience in Airbnb. We chose these dimensions to address issues of selection bias and treatment effect. Individuals self-select into their roles as either host or guest, and Airbnb has no way to capture the drivers of such decisions. As a consequence, it was important to separate hosts from guests in our analysis, allowing us to make the proper comparisons. Furthermore, we hypothesized that changes in the information used to trust strangers, as the result of a traveling experience with Airbnb, would be most pronounced for first-time users compared to experienced users. We thus divided level of prior experience into three tiers: (a) no prior experience, (b) moderate prior experience (two to five experiences as a host or guest), and (c) high level of prior experience (more than two to five experiences as a host or guest).

Finally, each cell of the stratified sample was divided into treatment and nontreatment groups. The treatment group was composed of Airbnb users scheduled to have a traveling experience within three weeks after Phase 1 of the study. The nontreatment group was made up of people not expected to travel during the same timeframe. The predictions about traveling were based on whether users had booked a room during the period of observation. These predictions were calculated by a team of researchers at Airbnb on our behalf. For each stratum of the random sample, treatment and nontreatment groups had roughly equal sizes and were roughly balanced on key dimensions.

Each participant was tagged with a unique token, namely, an alphanumeric label of 32 bits. The tokens were embedded in our invitation emails and served two purposes: (1) They increased the security of the study because only users with the tokens were allowed to register and play the game, and (2) they allowed us to match the experimental data collected in Phases 1 and 2 on the experiment website to the data Airbnb had on the same users. The mapping of experimental data with observational data is a key aspect of an online field experiment. As we will explain in greater detail in the following, this mapping is fundamental for preserving the anonymity of the participating subjects.

About 9,000 respondents accepted the email invitation originally sent by Airbnb and registered to our website, a response rate of about 8 percent. Because each individual in the sample was tagged with a unique token, we were able to collaborate with Airbnb to explore whether the individuals who responded to our invitation were systematically different from those who did not by examining Airbnb user data. We detected significant bias for both guests and hosts who accepted our invitation.

In particular, among the guests who had no prior experience, males were significantly less likely to have accepted our invitation than females. Furthermore, the gender bias was more pronounced among participants in the nontreatment group than in the treatment group. For guests with moderate prior experience, the gender bias remained, but the difference between the treatment and nontreatment groups was not as pronounced. Interestingly, these participants were younger than nonparticipants. Finally, participation in the study among experienced users had a bias similar to that of participants with no experience. The bias profile of the hosts who participated in the study was very different from that of the guests described previously. In general, hosts who participated had more ratings than hosts who did not participate. That is, the hosts included in our online field experiment tended to be more active than hosts in the larger sample.

APPLYING ONLINE FIELD EXPERIMENTS TO AN EXISTING STUDY

The central point of our argument is that online field experiments are becoming widely applicable because the social reality in which we live is ever more digitally mediated. This reality creates the possibility of incorporating greater levels of treatment complexity in the study of human behavior, where laboratory experiments do not allow us to isolate the effects of such complex treatments in which behavior is deeply environmentally embedded. In this section, we briefly illustrate how one *existing* study could apply an online field experiment design to account for the type of treatment complexity inherent in the theory being tested.

Burt's (2012) work on the role of agency in social networks uses data from a virtual world (Second Life and Ever Quest 2) to investigate the question of why individuals differ in the level at which they take advantage of structural holes in their networks. Structural holes occur between networks when those networks are not connected. Brokers can bridge structural holes so that the otherwise unconnected networks become connected through the broker. Burt (1992, 2004) theorizes that structural holes provide an information advantage for individuals in brokerage positions. Thus, an expectation of the theory is that brokers will be associated with more rewards and better individual-level performance than individuals embedded in more closed networks. Yet, considerable variation exists in the correlation between structural holes and the amount of rewards flowing to individuals. Burt highlights the role of agency in explaining this variation. Could it be that only individuals with certain personality traits are capable of taking advantage of structural holes? In his 2012 article, Burt tests the role of agency by observing how the same individual creates networks for multiple characters in virtual reality. The outcome measure is total points accumulated in the game at the end of the observation period. The key independent variable is the amount of brokerage in each network.

Using the online field experiment method, a researcher could engage with the online educational platform Coursera to directly test the effect of brokerage and agency on performance. The first step would be to create a random sample of users in a randomly selected number of courses and administer to them a personality test on the "self-monitoring" scale that Burt reports as important for understanding the connection between network creation and brokerage. Working directly with Coursera to glean network data from class discussion forums, one could use predictive modeling to identify users inclined to closure and brokerage (our treatment variable). The predictive model will use information collected on platform members, such as the text of their messages to other members, the length of engagement with the material, the quality of the assignment, and so forth. In addition to administering the personality test and any other survey measurements to these "participants," the researcher would measure participants' average and course-specific performance (our outcome variable) prior to the experiment. Through collaboration with Coursera, a researcher could observe changes in these participants' network structures over a period of time, such as the average time to completion of a Coursera course. The researcher would then measure the participant's performance again, noting any fluctuation in performance following brokerage or self-monitoring. In line with Burt's theory, self-monitoring would not be expected to correlate with course performance among participants with high network closure, self-monitoring would correlate highly with brokerage, and brokerage would be expected to result in stronger course performance.

This example represents one way in which online field experiments can be used to extend and replicate existing social psychological research. It also helps to build new bodies of information about emerging forms of social interaction and social organization in which interaction is heavily computer mediated, typically on the World Wide Web. Engaging with the platforms that supply access to such venues in which we can study the emergence of new social structures and their effects is an important methodological advance. This method extends the range of domains in which the development and testing of social psychological and sociological theories can occur. However, this promise does not come without clear challenges.

ETHICAL CONSIDERATIONS OF ONLINE FIELD EXPERIMENTS

The expanded use of online field experiments creates a new set of ethical concerns for researchers. Access to private information is at the core of this methodology because it allows linking real-life experience (i.e., the complex treatment) with measurement of its effect in the online "laboratory." Yet, users of the targeted online community may not want to give researchers access to their private information. Moreover, even if they were to agree to participate in a research project, they typically did not provide their information to the platform with research purposes in mind. How then can a researcher interested in using the online field experiment design address these ethical concerns?

Ethical design begins with the initial collaboration between the researcher and the online community platform. The goal of platform providers is to promote participation on their platform. Providers thus seek to avoid engaging in research activities that would abuse their users in any way. Just as in offline field studies, researcher interference in the online community can result in expulsion from the field site, among other negative outcomes. Ethical researchers should work carefully with the platform providers to ensure that recruitment and other forms of engagement with their participants are voluntary and nonintrusive. This entails outlining with providers potential harm to users before the study begins.

An important solution to the problems associated with ethical online field experiments is in using technology in ways that allow users to maintain control of their private information. Linking the treatment to measurement of its effects can be done using unique identifiers, for example, rather than private information such as names or email addresses. Unique identifiers can be tokens composed of a long sequence of letters and numbers that are assigned to each participant. The researcher can then use these tokens to link treatment conditions and the relevant measurements. The administrators of the platform will have access to the connecting table where private data and the associated tokens are stored, but the researchers would not. More importantly, the administrators of the platform could then completely anonymize the information they provide to researchers by sharing tokens rather than any private information.

To better illustrate this solution, we briefly describe the process we followed in our recent Airbnb study. After the company generated a random sample of users, we sent them a list of automatically generated tokens. Airbnb data scientists assigned each token to a user in our sample. They then sent emails on our behalf inviting users to participate. The emails contained a link to the website for the experiment where measurement of the complex treatment took place. The first page of the website showed a waiver of consent (that the university Institutional Review Board [IRB] approved for the purpose of this research). More importantly, each email was personalized with the token assigned to that specific user. If the user agreed to participate in the study, he or she would register, and we would then collect the unique token.

For our project, we also collected email addresses provided by the participants. However, depending on your collaboration with the platform, even this step is not necessary. We could have only relied on the tokens to join the data collected on our website with the data collected by Airbnb on the treatment. We opted to collect private information because we decided to send reminder emails to participants in between study waves to increase participation. However, Airbnb could have decided to send these reminder emails instead of us. The strength of this design is that it leaves the users of the platform, Airbnb in this case, in full control of the private information. While privacy concerns are paramount in online field experiments in which user data is needed to test theory, such concerns are often quite manageable with proper protocols that are now being developed for such research, often by internal IRBs within the online community platform and in collaboration with university-based IRBs.

CONCLUSION

In a world in which the traces of social interactions are increasingly available online, we propose a research methodology that extends the elements of experimentation to include greater treatment complexity. The basic idea behind this hybrid variant of field and laboratory methods is that more social interactions are occurring through technologies that leave "digital footprints" for analysis in the new era of "Big Data" (Golder and Macy 2014). The Internet is thus not just another mechanism for recruiting subjects but rather a new space where interactions occur and acquire meaning.

For experimentalists, this is good news. It means that greater levels of treatment complexity can now be more rigorously studied. Complex treatments are those that cannot be replicated in a lab because their meaning is fundamentally intertwined with the environment in which they occur. In this article, we provide an example of such a complex treatment: engagement with strangers in a sharing economy community. This treatment is patently not replicable in a laboratory because of the variation in the interactions that compose the sharing economy and because these interactions acquire meaning in the contexts in which they occur—a guestroom in a foreign country, a ride home late at night, or even a homemade meal delivered to your door.

The online field experiment methodology provides a method for incorporating the complexity of such treatments into a more traditional experimental design. More importantly, this methodology advances the argument that we are fundamentally living in a new social world—a world where interactions are increasingly mediated by technology. Investigation of these new and highly prevalent forms of social interaction and social structure calls for new methods, new concepts, and new analysis techniques.

ACKNOWLEDGMENTS

We are indebted to Xiaolu Wang and Morris Zelditch for their helpful suggestions.

FUNDING

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research is funded by the National Science Foundation (Grant #1257138).

REFERENCES

Asch, Solomon E. 1951. "Effects of Group Pressure upon the Modification and Distortion of Judgement." Pp. 177–90 in *Groups, Leadership and Men*, edited by H. Guetzkow. Pittsburgh, PA: Carnegie Press.

Ashmore, Richard D., Kay Deaux, and Tracy McLaughlin-Volpe. 2004. "An Organizing Framework for Collective Identity: Articulation and Significance of Multidimensionality." *Psychological Bulletin* 130(1):80–114.

Baldassarri, Delia, and Guy Grossman. 2013. "The Effect of Group Attachment and Social Position on Prosocial Behavior. Evidence from Lab-in-the-Field Experiments." *PLoS ONE* 8(3):e58750.

Berg, Joyce, John Dickhaut, and Kevin McCabe. 1995. "Trust, Reciprocity, and Social History." *Games and Economic Behavior* 10(1):122–42.

Bond, Robert M., Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler. 2012. "A 61-Million-Person Experiment in Social Influence and Political Mobilization." *Nature* 489(7415):295–98.

Burt, Ronald. 1992. *Structural Holes*. Boston, MA: Harvard University Press.

Burt, Ronald. 2004. "Structural Holes and Good Ideas." *American Journal of Sociology* 110(2):349–99.

Burt, Ronald. 2012. "Network-Related Personality and the Agency Question: Multirole Evidence from a Virtual World." *American Journal of Sociology* 118(3):543–91.

Campbell, Donald T., and Julian Stanley. 1963. *Experimental and Quasi-Experimental Designs for Research*. Chicago, IL: Rand-McNally.

Centola, Damon. 2010. "The Spread of Behavior in an Online Social Network Experiment." *Science* 329(5996):1194–97.

Chen, Yan, and Joseph Konstan. 2015. "Online Field Experiments: A Selective Survey of Methods." *Journal of the Economic Science Association* 1(1):29–42.

Cook, Karen S. 2005. "Networks, Norms and Trust: The Social Psychology of Social Capital." *Social Psychology Quarterly* 68(1):4–14.

Cook, Thomas D., and Donald T. Campbell. 1979. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Boston, MA: Houghton Mifflin Company.

Correll, Shelley J., Stephen Benard, S., and In Paik. 2007. "Getting a Job: Is There a Motherhood Penalty?" *American Journal of Sociology* 112(5):1297–339.

Darley, John M., and B. Latane. 1968. "Bystander Intervention in Emergencies: Diffusion of Responsibility." *Journal of Personality and Social Psychology* 8(4):377–83.

DeVoe, Sanford E., and Jeffrey Pfeffer. 2007. "Hourly Payment and Volunteering: The Effect of Organizational Practices on Decisions about Time Use." *Academy of Management Journal* 50(4):783–98.

Facebook Company Info. 2016. Retrieved April 2, 2016 (http://newsroom.fb.com/company-info).

Golder, Scott A., and Michael W. Macy. 2014. "Digital Footprints: Opportunities and Challenges for Online Social Research." *Annual Review of Sociology* 40:129–52.

Goldstein, Noah J., and Nicholas A. Hays. 2011. "Illusory Power Transference: The Vicarious Experience of Power." *Administrative Science Quarterly* 56(4):593–621.

Hahl, Oliver, and Ezra W. Zuckerman. 2014. "The Denigration of Heroes? How the Status Attainment Process Shapes Attributions of Considerateness and Authenticity." *American Journal of Sociology* 120(2):504–54.

Hauben, Michael, and Ronda Hauben. 1998. "Cybernetics, Time-Sharing, Human-Computer Symbiosis and Online Communities: Creating a Supercommunity of Online Communities." *First Monday* 3(8).

Holt, Charles A., and Susan K. Laury. 2002. "Risk Aversion and Incentive Effects." *American Economic Review* 92(5):1644–55.

Kohavi, Ron, and Roger Longbotham. Forthcoming. "Online Controlled Experiments and A/B Tests." In *Encyclopedia of Machine Learning and Data Mining*, edited by C. Sammut and G. Webb. New York: Springer.
Latane, Bibb, and J. M. Darley. 1968. "Group Inhibition of Bystander Intervention in Emergencies." *Journal of Personality and Social Psychology* 10(3):215–21.
McFarland, Daniel A., Kevin Lewis, and Amir Goldberg. 2015. "Sociology in the Era of Big Data: The Ascent of Forensic Social Science." *The American Sociologist* 47(1):12–35.
Ong, David, and Jue Wang. 2015. "Income Attraction: An Online Dating Field Experiment." *Journal of Economic Behavior and Organization* 111:13–22.
Owyang, Jeremiah, Alexandra Samuel, and Andrew Grenville. 2014. *Sharing Is the New Buying: How to Win in the Collaborative Economy*. Vancouver, BC, Canada: Vision Critical.
Page, Scott. 2015. "What Sociologists Should Know about Complexity." *Annual Review of Sociology* 41:21–41.
Pager, Devah, Bruce Western, and Bart Bonikowski. 2009. "Discrimination in a Low-Wage Labor Market: A Field Experiment." *American Sociological Review* 74(5):777–99.
PricewaterhouseCoopers. 2015. *The Sharing Economy*. New York, NY: PricewaterhouseCoopers.
Salganik, Matthew J., Peter Sheridan Dodds, and Duncan J. Watts. 2006. "Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market." *Science* 311(5762):854–56.
Santana, J., Raine Hoover, and Meera Vengadasubbu. 2017. "Investor Commitment to Serial Entrepreneurs: A Multilayer Network Analysis." *Social Networks* 48:256–69.
Santana, J., and Paolo Parigi. 2015. "Risk Aversion and Engagement in the Sharing Economy." *Games* 6:560–73.
Shepherd, Hana, and Elizabeth Levy Paluck. 2015. "Stopping the Drama: Gendered Influence in a Network Field Experiment." *Social Psychology Quarterly* 78(2):173–93.
Visual.ly. 2012. "The Numbers Behind the Sharing Economy. Retrieved November 16, 2016 (http://visual.ly/numbers-behind-sharing-economy).
Wiener, Norbert. 1948. *Cybernetics*. New York, NY: John Wiley & Sons.
Webster, Murray, Jr., and Jane Sell. 2014. "Why Do Experiments?" Pp. 5–21 in *Laboratory Experiments in the Social Science*, edited by M. Webster, Jr. and J. Sell. Boston, MA: Academic Press.
Zelditch, Morris, Jr. 1969. "Can You Really Study an Army in the Laboratory?" Pp. 528–39 in *A Sociological Reader on Complex Organizations*, edited by A. Etzioni. New York, NY: Holt, Rinehart and Winston.
Zelditch, Morris, Jr. 2014. "Laboratory Experiments in Sociology." Pp. 87–112 in *Laboratory Experiments in the Social Science*, edited by M. Webster, Jr. and J. Sell. Boston, MA: Academic Press.

BIOS

Paolo Parigi is the Lead Trust Scientist at Uber and Adjunct Professor of Civil and Environmental Engineering at Stanford University. He is interested in the relationship between risk and trust in the sharing economy. In 2013, Paolo received funding from the National Science Foundation to lead research in this area. With an interdisciplinary team of social scientists and computer scientists, Paolo has developed novel methodological approaches to harness the power of online data and has explored innovative ways for collaborating between academia and industry. His work may be relevant outside of the sharing economy to other areas in which technology is penetrating and reshaping human interactions. Paolo has an MA in quantitative methods and a PhD in sociology from Columbia University.

Jessica J. Santana is a doctoral candidate in the Department of Sociology at Stanford University. Her research evaluates the role of networks in innovation. This work is driven by insights from organizational theory, social psychology, network science, and economic sociology. She relies on a variety of methodological approaches, including experimental, statistical, and computational analyses. She has benefited greatly from experience working in the types of organizations she studies, from Silicon Valley startups to Nicaraguan farming cooperatives.

Karen S. Cook is the Ray Lyman Wilbur Professor of Sociology, director of the Institute for Research in the Social Sciences, and vice-provost for faculty development and diversity at Stanford University. She conducts research on social interaction, social networks, social exchange, and trust. She has edited a number of books in the Russell Sage Foundation Trust Series, including *Trust in Society* (2001), *Trust and Distrust in Organizations: Emerging Perspectives* (with R. Kramer, 2004), *eTrust: Forming Relations in the Online World* (with C. Snijders, V. Buskens, and C. Cheshire, 2009), and *Whom Can You Trust?* (with M. Levi and R. Hardin, 2009). She is coauthor of *Cooperation without Trust?* (with R. Hardin and M. Levi, 2005), and she coedited *Sociological Perspectives on Social Psychology* (with G. A. Fine and J. S. House, 1995). In 1996, she was elected to the American Academy of Arts and Sciences and in 2007 to the National Academy of Sciences. In 2004, she received the ASA Social Psychology Section Cooley Mead Award for Career Contributions to Social Psychology.